

Reinterpreting the development of reading skills

SCOTT G. PARIS

University of Michigan, Ann Arbor, USA

Learning to read is one of the greatest accomplishments in childhood because it is the foundation for learning and academic achievement. Therefore, it is not surprising that debates among educators about how best to help children learn to read have been heated, polarized, and unsettled for many years. The intensity of the debates, coupled with enormous political pressures and commercial interests, has made learning to read a contentious public issue in the United States. What makes the debates at the beginning of the new century different than similar debates during the past 50 years is a greater than ever reliance on scientific evidence to guide educational policies for assessment and instruction. The attention and new credibility given to reading research have been hard won by the academic community and have great promise, but there are also pitfalls to avoid in the rush to use basic research for legislated policies and educational prescriptions (Berliner, 2002; Feuer, Towne, & Shavelson, 2002).

The purpose of this article is to reveal flaws in traditional research on developing reading skills that have wide ramifications. My goal is to stimulate researchers to reconsider how to conceptualize, research, and interpret the development of reading skills. This reinterpretation is necessary because traditional reading research has ignored fundamental differences in the developmental trajectories of reading skills. These different trajectories are manifested in different times of skill onset, different durations of acquisition, and different asymptotic levels of performance. What is most important is that there are some skills that are more constrained than others; they are learned quickly, mastered entirely, and should not be conceptualized as enduring individual difference variables. Constrained skills have been analyzed with the same research tools and parametric statistical analyses as unconstrained skills. I claim that this is a mistake that can lead to spurious claims about early reading skills. The theoretical focus of this article is on construct validity and the developmental differences among reading skills. The methodological focus is on the distributions of longitudinal data and the methods to examine

THEORIES ABOUT reading have neglected basic differences in the developmental trajectories of skills related to reading. This essay proposes that some reading skills, such as learning the letters of the alphabet, are constrained to small sets of knowledge that are mastered in relatively brief periods of development. In contrast, other skills, such as vocabulary, are unconstrained by the knowledge to be acquired or the duration of learning. The conceptual, developmental, and methodological constraints on different reading skills are described in this essay that identifies various types of constraints on reading constructs and measures. Examples of reading research and assessment are discussed to illustrate (a) how the constraints can help to explain transitory correlational patterns among reading data, (b) how proxy effects surrounding constrained skills influence interpretations of reading development, (c) how prescriptions to teach constrained skills are causal misinterpretations of longitudinal correlations, and (d) why interventions on constrained skills usually lead only to temporary gains on skills aligned with the constrained skill. Because constrained skills are not normally distributed conceptually or empirically, except on special occasions, analyses based on parametric statistics are inappropriate. This essay describes implications for theories of reading development, research methods, and educational policies.

LAS TEORÍAS acerca de la lectura han descuidado diferencias básicas en el camino evolutivo de las habilidades relacionadas con la lectura. Este ensayo propone que ciertas habilidades de lectura, como el aprendizaje de las letras del alfabeto, están restringidas a pequeños conjuntos de conocimientos cuyo dominio se concreta en períodos relativamente breves. En contraste, otras habilidades tales como el vocabulario, no están restringidas por el conocimiento que se adquirirá ni por la duración del aprendizaje. En este ensayo se describen las restricciones conceptuales, evolutivas y metodológicas en diferentes habilidades de lectura; el mismo identifica varios tipos de restricciones en las conceptualizaciones sobre la lectura y en las medidas de lectura. Se discuten ejemplos de investigación y evaluación en lectura para ilustrar: a) de qué modo las restricciones pueden ayudar a explicar patrones correlacionales transitorios entre los datos de lectura, b) cómo efectos secundarios próximos a las habilidades restringidas influyen las interpretaciones del desarrollo en lectura, c) cómo ciertas propuestas para enseñar habilidades restringidas son malas interpretaciones causales de correlaciones longitudinales y d) por qué las intervenciones sobre las habilidades restringidas generalmente conducen sólo a logros temporarios en habilidades afines a la habilidad restringida. Debido a que las habilidades restringidas, excepto en ocasiones especiales, no están normalmente distribuidas conceptual o empíricamente, son inapropiados los análisis basados en la estadística paramétrica. El ensayo describe implicancias para las teorías del desarrollo lector, los métodos de investigación y las políticas educativas.

THEORIEN ÜBERS Lesen haben die grundsätzlichen Unterschiede in den aufs Lesen bezogenen sich entwickelnden Leistungsbahnen vernachlässigt. Dieser Aufsatz vertritt die Ansicht, daß einige Leseleistungen, wie beispielsweise das Erlernen der Buchstaben des Alphabets, auf eng begrenzte Erkenntniswerte beschränkt sind, die in relativ kurzen Entwicklungsperioden beherrscht werden. Im Gegensatz dazu sind andere Kenntnisse, wie das Vokabular, unbegrenzt im zu erlernenden Wissen oder in der Dauer beim Lernen. Die konzeptuellen, entwicklungsbedingten und methodologischen Beschränkungen auf unterschiedliche Leseleistungen werden in diesem Aufsatz beschrieben, welcher verschiedene Typen von Einschränkungen im Lesen von Wortfügungen und Abschätzungen identifiziert. Beispiele aus der Leseforschung und Bewertung werden diskutiert, um zu illustrieren, (a) wie die Einschränkungen dazu verhelfen können, transitorische Korrelationsmuster zwischen den Lesedaten zu erklären, (b) wie Vertretungsfunktionen umgebende Auswertungen von Leistungseinwirkungen der Leseentwicklung beeinflusst, (c) wie Verordnungen zum Unterrichten eingegrenzter Kenntnisse kausale Fehlinterpretationen der Längskorrelationen sind, und (d) warum Interventionen bei eingegrenzten Kenntnissen meistens nur zu zeitweisen Fortschritten der angeglichenen Leistungen eingeschränkter Fähigkeiten führen. Da eingeschränkte Befähigungen normalerweise nicht konzeptionell oder empirisch eingeteilt sind, außer bei speziellen Anlässen, sind die auf parametrische Statistiken basierte Analysen ungeeignet. Dieser Aufsatz beschreibt Implikationen für Theorien zur Leseentwicklung, Forschungsmethoden und Unterrichtsverfahren.

ABSTRACTS

Reinterpreting
the development
of reading skills

Reinterpretando
el desarrollo de
las habilidades
de lectura

Reinterpretieren
der Entwicklung
von Leseleistungen

読みの技術の発達 の再解釈

読みの理論は、読みに関連する技術の発達上の軌道における基本的な違いを無視してきた。本稿では、アルファベットの文字の学習のような読みの技術は、比較的短い発達 の時期に身に付く小知識に制約されていると提唱する。対照的に、語彙のような他の技術は、習得される知識や学習の期間によって制約されない。様々な読みの技術に対する概念的、発達の、方法論的制約が、読みの構成概念及び測定数に対する様々な制約を識別する本稿において叙述される。(a) どのようにそうした制約が読みのデータの一時的な相関パターンを説明するのに役立つのか、(b) どのように制約された技術を取り囲む代理効果が読みの発達の解釈に影響を与えているのか、

(c) どのように制約された技術を教えるための規定が、縦断的相関の因果関係の間違った解釈であるのか、(d) なぜ制約された技術への介入が、大抵の場合、制約された技術と連結した技術の一時的な伸びにしか至らないのかを説明するために、読みの研究と評価の例を論じる。制約された技術は、特別な場合を除いて、概念的にも経験的にも正規分布されないため、パラメトリック検定による分析は不適当である。本稿は、読みの発達理論、研究方法、そして教育方針のための示唆を述べる。

Réinterprétation
du développement
des compétences
de lecture

LES THÉORIES de la lecture ont négligé des différences de base dans les trajectoires du développement des compétences relatives à la lecture. Cet essai énonce que certaines compétences de lecture telles que les lettres de l'alphabet se réduisent à de petits ensembles de connaissances qui sont maîtrisées au cours de périodes de développement relativement brèves. Par contre, d'autres compétences telles que le vocabulaire ne sont pas limitées par les connaissances à acquérir ou par la durée de l'apprentissage. On décrit dans cet essai, qui identifie différents types de contraintes sur les constructs et les évaluations de la lecture, les contraintes conceptuelles, développementales, et méthodologiques de différentes compétences de lecture. On discute des exemples de recherche et d'évaluation de la lecture afin de montrer a) comment les contraintes peuvent aider à expliquer certains patrons provisoires de corrélations entre résultats de lecture, b) comment des effets voisins qui se trouvent autour des compétences réduites jouent un rôle sur les interprétations du développement de la lecture, c) comment les prescriptions d'enseignement de compétences réduites sont responsables d'erreurs d'interprétation des corrélations longitudinales, et d) pourquoi des interventions sur des compétences réduites ne conduisent en général qu'à des gains temporaires sur des compétences liées à celles-ci. Du fait que, sauf en de rares occasions, les compétences réduites ne sont pas distribuées normalement, tant conceptuellement qu'empiriquement, les analyses reposant sur des statistiques paramétriques sont inappropriées. Cet essai énonce enfin des implications sur les théories du développement de la lecture, les méthodes de recherche, et les politiques éducatives.

Иное толкование
развития
навыков чтения

Теории чтения пренебрегают тем фактом, что индивидуальные траектории развития навыков, тем или иным образом связанных с чтением, весьма различны. Некоторые из этих навыков, например, освоение букв алфавита, автор предлагает преподавать компактно, в виде достаточно ограниченного блока знаний, освоить которые предлагается за относительно краткий срок. Напротив, другие навыки, такие как, к примеру, использование и расширение словарного запаса, не должны быть жестко вписаны в определенный блок знаний или единицу времени, отведенную для их освоения. В статье классифицированы концептуальные, методические и индивидуальные параметры и ограничения, влияющие на развитие навыков чтения. Приводятся примеры исследований самого чтения и его оценивания, чтобы проиллюстрировать, (a) что знание о принудительном развитии навыков может помочь в выявлении и трактовке временных и непрочных результатов, (б) что промежуточные результаты, полученные в результате принудительного развития навыков, влияют на интерпретацию полученных данных, (в) что рекомендации по принудительному развитию навыков возникают в результате неверного истолкования кратковременных корреляций как долговременных и (г) что принудительное развитие навыков обычно ведет только к временным достижениям во всем, что связано с этими навыками. Поскольку связанные навыки крайне редко развиваются концептуально или эмпирически, анализ, основанный на параметрической статистике здесь неприемлем. Предлагаются выводы, связанные с теорией развития навыков чтения, методов исследования чтения и образовательной политикой.

relations among developing skills. The interpretive focus is on differences between prediction and causation and the implications for assessment and instruction practices. Because the scope of this article is broad and the implications for researchers and policymakers are far-reaching, the article includes both representative empirical studies and conceptual generalizations. I hope that it begins a constructive debate on new approaches to study reading development.

Rationale for reconceptualizing reading skills

Many theories of reading (e.g., Chall, 1967, 1996; Clay, 1991; Ehri, 1995; LaBerge & Samuels, 1974; Rumelhart, 1994) propose that multiple skills are learned during childhood, at home and in school, and they become coordinated into increasingly automatic reading in grades 1–5. The theories regard skills as components to be acquired and assembled, and the main controversies in the so-called “reading wars” have been arguments about the developmental order and importance of decoding versus comprehension skills. I want to avoid a discussion of the controversy, but I will note that the current compromise emphasizes a balance of both kinds of skills in early instruction and a recognition that they interact and may compensate for each other during reading (Adams, 1990; Stanovich, 1980). This uneasy compromise has not changed the prevailing views of reading development as the compilation and coordination of many component skills or led to more detailed examination of the differences among reading skills (e.g., National Institute of Child Health and Human Development [NICHD], 2000).

One problem with the general notion of assembly and automatic use of component skills is that all skills are regarded as similar in scope, importance, and enduring individual differences. The National Reading Panel report (NICHD, 2000) and *Preventing Reading Difficulties* edited by Snow, Burns, and Griffin (1998) identified five essential component skills for reading development: the alphabetic principle, phonemic awareness, oral reading fluency, vocabulary, and comprehension. These skills, and the implicit developmental model of assembly and automation, in turn became the foundation for the Reading First legislation that was part of the No Child Left Behind Act (2002). That federal legislation has had enormous impact on classroom instruction and assessment of reading in grades K–3

across the United States. Specifically, there has been increased assessment and instruction on alphabet knowledge, phonemic awareness, and oral reading fluency as the main enabling skills and significant predictors of later reading achievement. There has been relatively less research and classroom emphasis on vocabulary and comprehension to date, perhaps because of the difficulty assessing and teaching these skills to children who are beginning to read.

Despite an abundance of research to support policies based on the five essential skills, I suggest that the underlying research should be reexamined because alphabetic knowledge, phonemic awareness, and oral reading fluency are constrained both theoretically and methodologically, unlike vocabulary and comprehension. They all develop from nonexistent to high or ceiling levels in childhood. One consequence of these developmental trajectories of mastery is that the distributions of data are highly variable and unstable over time for constrained skills. They are usually and necessarily skewed during initial acquisition and later mastery with variance that ranges from nil to large to nil during mastery. All statistical variances and correlations are unstable longitudinally. Thus, parametric statistics, such as Pearson correlations and ANOVAs, are inappropriate for analyzing data derived from constrained skills. This is not simply a problem with skewed data that can be repaired with statistical transformations; it is a challenge to understand that the constructs are defined and operationalized in ways that make normal distributions of data unlikely. Unraveling the confounds among reading skills reveals how different they are in developmental trajectories of acquisition, why traditional correlational analyses may not be appropriate, and why traditional research on the five essential skills may need to be reinterpreted.

Constraints on reading skills

For descriptive convenience, the constraints that influence analyses of reading development can be grouped in three categories: conceptual, developmental, and methodological. The type and degree of constraints can vary widely among reading skills. In general, letter knowledge, phonics, and concepts of print are highly constrained, phonemic awareness and oral reading fluency are less constrained, and vocabulary and comprehension are least constrained. The polarized ends of the continuum exemplify the differences between constrained and unconstrained skills most clearly, but even subtle constraints hinder analyses of reading skills.

Conceptual constraints

Reading skills, like other psychological constructs, are defined and operationalized in order to establish consistent measurement and interpretation of the constructs. The validity of the construct is evaluated by reference to its definitions and measures. One fundamental constraint of a construct is the *scope*, defined by its domain, number of elements, or set size. For example, some skills and concepts are constrained by the number of elements that must be acquired. Learning the names and sounds of the 26 letters in the English alphabet, usually referred to as letter knowledge and phonics, are two clear examples. Learning letter-sound relations and identifying both lower- and uppercase letters are constrained skills because the number of elements to be mastered is small and finite. (Languages around the world may vary in the numbers of sounds and symbols they contain, but they all are finite sets that are mastered by literacy users.) Skills with narrow scope are learned quickly so the trajectory of mastery is steep and the duration of acquisition is brief. The same is true of basic concepts about print, such as those described by Clay (1979) in her observation survey, which have been used and adapted in many early reading assessments. These can include concepts about word boundaries, sentences, punctuation marks, directionality of reading, and other features of text orthography. They are conceptually constrained by the relatively small number of concepts to be acquired.

Some might object that concepts about print are really without limits, and they might give examples of genre-based rules that are seldom acquired by many adults to show that these are unconstrained skills. There is some merit to that argument theoretically, but in practice the number of print concepts that beginning literacy users acquire is limited to the most important and prototypical concepts, and those are relatively few. Thus, the second conceptual constraint on early reading skills and concepts is *importance* as measured by centrality or typicality of exemplars. Instruction and assessment of early reading both concentrate on the important foundational concepts that children need to acquire (e.g., word and sentence boundaries, top down and left to right direction of reading English, capital versus small letters, and meanings of punctuation marks). Likewise, early instruction and assessment of letter knowledge begins with frequently used letters with clear grapheme-phoneme correspondence. Thus, most assessments of children's understanding of early print concepts pertain to a small set of central and important features of text that beginning readers need to understand.

A third type of conceptual constraint among reading skills is the *range of influence*, both the domain and temporal range. The concomitant skills influenced by learning the alphabet and concepts of print are tied directly to decoding grapheme-phoneme relations. In contrast, vocabulary development influences (and is influenced by) linguistic, cognitive, and communicative proficiency in wide-ranging ways. Not only is the range of influence less among constrained skills, but also the temporal range of influence is restricted in constrained skills to that period when rapid acquisition is occurring. Thus, constrained skills such as alphabet knowledge are most related to decoding in early childhood, whereas unconstrained skills such as vocabulary are related to a wide range of academic skills throughout life.

Scope, importance, and range of influence are also evident in constructs involving phonological awareness, although each construct is broader than alphabet knowledge and acquisition takes longer. Most early reading assessments include tasks of phoneme identification, segmentation, blending, and rhyming, but the phonemes that are used in the tests are usually important and prototypical. They are based on phonemes and words that have a high frequency of occurrence in oral language and text, they are usually familiar, and they have a wide range of application. Thus, words such as *cat* and *sit* are frequent in early assessments of phonological awareness because the onset-rime patterns are central prototypes for young children. Because most assessments of phonological awareness are limited to a small number of rules that are assessed with central exemplars, they can be considered constrained skills.

Although it is possible to create assessments of infrequent and obscure phonological patterns and thus derive assessments that are unconstrained by the numbers of rules and elements as well as age and duration of acquisition, this is not done in practice or research because it is not relevant to children's beginning reading. Nearly all assessments of children's phonological awareness concentrate on a few central rules exactly because the mastery of these rules is fundamental. Those rules typically include onset-rime patterns, initial consonants, final consonants, vowel patterns, segmentation, and blending that are appropriate for children in grades K-2 (e.g., *The Fox in the Box* by Adams & Treadway, 2000; Texas Primary Reading Inventory [TRRI], 2001). Although there are many distinct phonological rules, and the set size is greater than 26, it is the constrained scope, importance, and range of influence of these rules that enable 7- and 8-year-old children to learn the essential features of phonological awareness and to

decode nearly all words they encounter in beginning texts.

Phonological processing is more complicated than phonological awareness because constrained and unconstrained skills may operate together. Thus, assessments of phonological processing may confound constrained and unconstrained skills. For example, the Comprehensive Test of Phonological Processing (CTOPP) by Wagner, Torgeson, and Rashotte (1999) is based on the interactions among three correlated skills: phonological awareness, phonological memory, and rapid naming that, according to the authors of the test, become less correlated with development. The constrained skills view provides a useful framework for understanding the CTOPP because phonological memory and rapid naming are both individual difference variables between children that endure over time, whereas phonological awareness is a constrained skill that varies within and between individuals during a period of mastery only. After mastery is achieved, phonological awareness carries little variance compared to phonological memory and rapid naming, so the correlations among the dependent skills must decrease for statistical reasons alone. What remains are individual differences in basic information processing functions of memory and rapid naming, not skills specific to reading.

The CTOPP includes 13 subtests, 7 for children ages 5–6 years and 10 for people 7–24 years old. The older subjects actually receive all of the same subtests as the younger sample except Sound Matching, presumably because it is mastered by age 7, and then 4 more difficult subtests, Blending and Segmenting Nonwords and Rapid Naming of Colors and Objects. The subtest scores are used to create three composite scores for Phonological Awareness, Phonological Memory, and Rapid Naming that are used to establish normative ages for performance. The composite scores are also used as correlates to establish concurrent and predictive validity. CTOPP combines very different skills in the composite measures that confound developmental and methodological constraints, and thus the interpretations of the data may have less to do with phonological awareness than individual differences in processing speed and memory.

Examination of Table C.1 in the CTOPP manual reveals wide variation in the developmental growth rates of the performance on the subtests between ages 5–15. For example, Sound Matching is at 100% by 8 years, 3 months; Blending Words is at 87% mastery by age 8 years, 9 months; Blending Nonwords is at 80% mastery by the same age; and Phoneme Reversal is at 70% mastery by age 9. Floor

effects are evident for Segmenting Words because all children who segment fewer than nine words receive the same low score. Growth from 0 to 9 on that scale is from ages 5 years to 7 years, 9 months, and increases in Segmenting Words only grows from a raw score of 9 to 12 between ages 7 years, 9 months and 14 years, 9 months. Clearly, growth on the subtests is not linear or uniform; ceiling and floor effects are evident in individual subtests so composite scores derived from the subtests reflect proficiency (or lack of it) for different skills at different ages. Consider Memory for Digits. The scale of raw scores is from 0 to 14 between ages 5–9 and 14–9, but the conversions to age equivalents show that 0 to 10 scores are all expected before age 6 with growth during the next 9 years limited to increases of four items. Failure to consider the constraints on various skills and different developmental trajectories of the skills undermines the construct validity of the composite scores in the CTOPP.

Details about CTOPP are presented here to provide a specific example of the differences in developmental trajectories and the asymptotes reached in assessment instruments. Similar criticisms can be directed to other early reading assessments including the Dynamic Indicators of Basic Literacy Skills, the Texas Primary Reading Inventory, and the Phonological Awareness Survey because they include assessments of constrained skills such as letter knowledge, concepts of print, and phonemic awareness. It is worth noting that these early skill assessments typically report Pearson correlations to show the relations with other reading tests to establish concurrent and predictive validity. However, the psychometric data are necessarily skewed from novice to expert status, and those variable distributions influence the correlations. That is why reconsideration of the differences among skills is important.

Developmental constraints

There are four important constraints on the developmental trajectories of reading skills. The first constraint is *unequal learning* because some letters, concepts, and phonemes are learned more quickly and thoroughly than others. They are fixed effects with heterogeneous variance that result in nonlinear learning among elements. For example, the letters *x* and *q* are learned later and more slowly than the letters *m* and *s*, and phoneme rhyming is generally easier with consonant-vowel-consonant words than more complex patterns. Unequal learning of exemplars of concepts or instances of rules becomes a problem when skills such as alphabet knowledge and phonemic

awareness are treated as uniform skills, especially in assessments that presume random sampling of equivalent elements.

A second developmental constraint is *mastery*. Some reading skills, such as learning the alphabet, are mastered completely, whereas other skills, such as vocabulary, are not. Whether the learning occurs during childhood or during adulthood does not change the fact that the degree of learning is complete. Moreover, the duration of learning of mastered skills is relatively brief. These temporal constraints are not evident in unconstrained skills that continue to develop over the life course. Mastered skills must exhibit floor and ceiling effects in the longitudinal course of acquisition because constrained skills develop from nonexistent to fully acquired to automatic. Granted that some reading skills may not be mastered perfectly or completely, they approach an individual growth asymptote as acquisition slows or a ceiling is attained. It seems likely that mastered skills for an individual follow a sigmoid growth function—an S-shaped curve—in which initial acquisition of a skill is slow, followed by a period of rapid learning, and then followed by a slower rate of growth as asymptotic performance is approached.

Walking and knowing the sounds of a language are good examples of mastered skills. Because performance on mastered skills is at similar levels for children of similar ages (e.g., walking), they are sometimes regarded as age-normative skills. In contrast, unconstrained skills continue to develop throughout the life span, are not identical across people, and may benefit from special practice and idiosyncratic experiences at many points in the life course. Manual dexterity and vocabulary are good examples of unconstrained (or at least much less constrained) skills. Although it is possible to calculate and display developmental norms for the growth of both constrained and unconstrained skills, unconstrained skills may exhibit more variation among people in the specific elements that are learned, the degree of expertise attained, and the onset and duration of acquisition. Constrained skills, because they are smaller sets of identical information, appear to be acquired in more uniform ways and rates, but uniformity is not necessary for individual mastery. It is also important to note that normative growth patterns of mastered skills do not imply that the underlying variables are normally distributed among people.

The third developmental constraint is *universality*. Some reading skills and concepts reveal mastery of identical information among people. Most skilled readers of English know the 26 letters of the alphabet and the phonemes associated with them.

Likewise, all competent readers know the identical (or nearly so) concepts about print and understand phonemic rhyming, segmentation, and blending in the same manner. On assessments of these reading skills, they would have the identical y-intercepts or asymptotes. This is a critical feature of constrained skills because it results in zero (or at least minimal) variance between individuals when the constrained skill is at asymptotic levels. Contrast a universal constraint with less constrained variables such as growth in height or vocabulary. Both may reach asymptote in adulthood, or at least exhibit slower rates of change, but the asymptotes are different across individuals and the differences in the y-intercepts are normally distributed. This is not the case for universally mastered skills that attain the identical intercepts and have no enduring individual differences. Thus, the differences during acquisition of universally mastered skills (in terms of onset, rate, or duration) are minor compared to the similarity over most of the life span. Unconstrained skills continue to develop over time and may reveal enduring differences between individuals over the life span. This is a crucial distinction that has implications for the kinds of statistical analyses and interpretations that are appropriate with each kind of skill.

Of course, universally mastered skills can vary between people, but the differences are only transitory. For example, 5-year-old Joe may know 18 letters of the alphabet while 6-year-old Mary may know 5 or 24 letters, but both children will know all 26 letters eventually. Constrained skills are distributed at different mastery levels between people only during the brief period of acquisition. They are mastered 100% by everyone eventually, whereas unconstrained skills are distributed between people on a norm-referenced continuum (e.g., the y-intercepts of height or vocabulary) over the life span. Unconstrained skills yield normally distributed variables between people over time, and if their ranks in the temporally spaced distributions remain similar, the differences are stable individual differences. However, constrained skills will approximate normal distributions only when the sample includes individuals who exhibit partial mastery. The actual distribution of data always reflects the specific mastery levels of the sample and is not a random or representative sample of the general population. Between-Ss differences in universally mastered skills are fleeting rather than stable individual differences. Research designs and statistical analyses that ignore this fact are incongruent with the underlying constructs of universally mastered skills.

The analysis of developmental trajectories is more complicated because the sigmoid growth

curves of various skills can vary widely in age of onset and duration of growth. Consider the subtests in the CTOPP. The normative data reveal that Sound Matching, Blending Words, and Phoneme Reversal are acquired rapidly by 7–8 years of age while Segmenting Nonwords and Elision are mastered more slowly. When they are combined in composite scores by statistical techniques, the fundamental differences among different growth curves are obscured. The data may have improved normality of dependent variables but at an undisclosed cost of reduced validity. Designers of the Woodcock-Johnson Reading subtest evidently recognized the problem with a low ceiling on letter naming so they created composite scores of letter naming and word identification. Creating composite scores to skirt skewed data solves the normality problem but confounds what is being measured.

There are many studies that illustrate how mastery of constrained skills may confound data analyses and interpretation. For example, Morris, Bloodgood, Lomax, and Perney (2003) reported a longitudinal study of the relations among reading skills in children during kindergarten and first grade. They used a LISREL model to test the relations among emerging skills, but they failed to consider the conceptual and methodological constraints in their data. For example, Table 3 in their report shows clearly that children's alphabet knowledge and beginning consonant awareness were at ceiling levels at Times 2, 3, and 4 in their study, while measures of children's word recognition and phoneme segmentation were at floor levels for Times 1, 2, and 3. Nevertheless, they used traditional statistical analyses that assume normally distributed measures. From their longitudinal analyses, they concluded that the developmental sequence of acquisition for seven reading skills is alphabet knowledge, beginning consonant awareness, concept of word in text, spelling with beginning and ending consonants, phoneme segmentation, word recognition, and contextual reading ability.

The latter two skills are unconstrained and would be predicted to develop longer and slower than the other five constrained skills. It is possible to predict the order of acquisition based only on the size of successively increasing knowledge sets within each skill and greater constraints in measurement among the early emerging skills. Morris et al. (2003) argued that concepts about words (i.e., finger pointing to words as the story is read) are important precursors to phonemic awareness and should be emphasized in instruction. This may be true only in the trivial sense that concepts about words is a small-

er set of knowledge that is learned more rapidly than phonemic awareness, and it is a necessary but not sufficient condition for other skills to emerge.

Although the empirical identification of patterns of emerging developmental skills is important, the interpretations would be more accurate and complete if the conceptual distinctions among developing skills were considered.

Other examples where constraints on skill development were ignored abound in recent publications. Thompkins and Binder (2003), for example, compared illiterate adults and children matched on reading levels on a variety of reading skills. Some of the skills, such as phoneme recognition, word-picture matching, and digit span were at ceiling levels that influenced the variance of the measures. Nevertheless, the researchers ignored both conceptual and methodological constraints on the variables and used multiple regression and ANOVAs to analyze the data. Their conclusions are thus open to reinterpretation. Likewise, Burgess, Hecht, and Lonigan (2002) studied the relations between home literacy environments and early reading skills of 4- and 5-year-olds in a one-year longitudinal study. Their tasks included phonological sensitivity tasks (e.g., rhyme oddity, blending of compound words, and an elision task), a letter-name knowledge task, a letter-sound knowledge task, and word decoding measures. Despite acknowledging that the data were not normally distributed and that transformations failed to correct the deviations, they analyzed the untransformed data with correlations and multiple regressions. The authors failed to consider how the relative degree of mastery of constrained skills can influence the correlations, and especially the influence on the autoregressor, and interpretations of the data. The conclusions are thus open to reinterpretation.

A fourth kind of developmental constraint on reading skills is *codependency*. Some precursors might be necessary for a skill to be acquired, so it is constrained by its relation to other skills. For example, there are many skills involved in language reception, discrimination, and production that underlie emerging literacy skills. These skills may be necessary prerequisites for literacy development. Specifically, phoneme identification of consonants precedes identification of vowels and may be a necessary precursor to segmentation and blending skills. In general, comprehension of text depends on decoding the words; decoding is a necessary but not sufficient condition for understanding text. Many constrained reading skills are dependent on cognitive and linguistic development and are acquired during childhood about the same time. The parallel and

simultaneous development of language and literacy skills leads to multicollinearity of these variables in research studies and makes it especially difficult to separate the relations among the skills during periods of rapid development. The codependency also may invalidate correlational analyses.

Researchers have analyzed reading skills as if they are independent when many are required as precursors or enabling skills for others. Thus, there is a positive correlation during acquisition between codependent skills that is logically necessary during acquisition, but the relation will disappear when both skills are fully mastered and there is minimal variance in either one. This pattern of transitory relations and later lack of relations between skills is a consequence of the developmentally codependent constraints on the skills. The constrained skills view explains the transitory correlation patterns as logical, conceptual, and empirical consequences of children's mastery of constrained skills as developmental precursors and enablers of other skills.

It is important to note that the codependency may be asymmetrical when one skill enables another. Thus, the lack of skill A may be correlated with the lack of skill B, if B depends on A, but the proficiency of skill A does not imply that skill B is also proficient (if A is a necessary but not sufficient condition for skill B). This asymmetrical relation is evident among novice readers when the lack of oral reading fluency is correlated with the lack of comprehension, but fluency is not necessarily correlated with comprehension among skilled readers (Paris, Carpenter, Paris, & Hamilton, in press). In more general terms, this means that emerging or novice skills may display codependent relations with other reading skills, as well as greater variance than the same subjects with highly proficient skills. Thus, positive correlations are evident only among partially developed skills and not evident at all among mastered skills. One implication of this asymmetrical relation is that positive correlations might only be observed for novice skill users, struggling readers, or readers with skill deficits. As a consequence, models of reading skills built only on struggling readers with poor skills may overemphasize asymmetrical relations between the lack of codependent skills. Deficit models of unskilled reading may provide an incomplete and inaccurate characterization of the developmental relations among reading skills.

Methodological constraints

Some constraints result from methods used to gather data. Mundane examples of measurement constraints include rubrics with narrow ranges and

unreliable interrater agreement and assessment tasks that are too easy or difficult so the data are skewed empirically but not necessarily conceptually. Codependency between skills provides a more compelling example of methodological constraints because the use of one skill may depend on a minimum or critical level of another skill. Consider two examples of reading skill "thresholds," oral reading accuracy (accuracy) and oral reading rate (rate). Both are constrained skills, yet they have been treated as unconstrained variables in traditional research. Accuracy is constrained conceptually because skilled reading is not distributed around a midpoint of 50% accurate word identification. Indeed, 100% accuracy is the goal and preferred skill level so it is constrained conceptually. Accuracy is also constrained as a research variable because educators and researchers consider a level of 95% accurate word identification to be essential for comprehending text (Lipson & Wixson, 2003). This means that accuracy (a) sets a threshold for comprehension and (b) is a highly skewed skill with limited variance when compared to other reading skills.

Reading rate is constrained by speed of speech production and automatic word recognition, but in practice is less constrained than accuracy. Most children read aloud at similar rates as they learn to decode words, and few read less than 40-50 words correct per minute (wcpm). The midrange of oral reading rate for first graders in the fall is about 53 wcpm, whereas fifth-grade students at the 50th percentile in the fall read about 105 wcpm (Hasbrouck & Tindal, 1992). At each successive grade level, children read on average about 13 more words per minute. Thus, the range of reading rate is constrained within and between grades by a modest range of growth each year. Reading very slowly or very quickly may degrade comprehension so oral reading rate is constrained by speech rate, expertise, and attempts to understand text while decoding print. Certainly the complexity of text, the familiarity of vocabulary, the audience, and the purpose for reading also influence rate of reading. The thresholds for reading accuracy and rate increase with age and instruction, but reading accuracy, and rate to a lesser degree, are constrained skills conceptually, developmentally, and empirically.

Correlations between fluency and comprehension for highly accurate oral readers have little variance to begin with (in the data), so modest correlations are the most that can be expected. Floor effects, though, are paradoxical because of the asymmetrical codependency. Even though there is little variance in very low scores on fluency and compre-

hension, they will covary by necessity when assessed on the same text because fluency enables comprehension (i.e., there can be no comprehension if the words cannot be read accurately). It makes no sense to assess the relation between fluency and comprehension when oral reading accuracy is at low levels because the relation will always reveal the obvious and spurious positive correlations. It should not be surprising that children who cannot recognize many words in a passage also cannot comprehend it. The nonindependence of the variables at low levels of decoding certainly confounds and inflates the positive relation, and it may invalidate correlational analyses involving oral reading accuracy.

So, why do some studies find modest positive correlations between oral reading fluency and comprehension (Kuhn & Stahl, 2003)? Sometimes the studies include data from many readers who are reading below 90% accuracy, so the data include cases of readers who cannot decode or comprehend the text. The variance in scores between 0 and 90% accuracy is huge and yields positive correlations with comprehension because both scores are so low for poor readers (see Kibby, 1979). Even when the majority of subjects have accuracy scores above 90%, the correlations with comprehension are unduly influenced by the few cases with the most variance, outliers in a statistical sense, because there is little variance in fluency scores among the best readers. Readers who have less than 90% accurate oral reading may have a wide variety of reading problems, including inadequate prior knowledge, poor vocabulary, unfamiliarity with standard English, unfamiliarity with the passage genre and test format, and motivational obstacles such as low self-efficacy and self-handicapping strategies (Paris & Paris, 2001). Thus, oral reading accuracy can be influenced by many different experiences and skills, and the oral reading fluency score may only be a proxy measure for many other influences on reading development.

Some researchers have correlated fluency and comprehension scores on different tasks to avoid the codependency problem between skills. For example, oral reading fluency can be calculated on one text and comprehension can be assessed on standardized reading tests. This seems questionable because the relation between the two cognitive processes is important within texts, not between texts. Assessing the skills on independent texts treats the skills as independent abilities so one can argue about which is the legitimate way to conceptualize the relation between fluency and comprehension. Correlations between fluency scores on one reading task and comprehension scores on another task generally yield positive

correlations. However, the interpretation is debatable. Advocates argue that the two skills are positively related because of the correlations, but the method of analyses on independent texts means that the relation is between subjects and not cognitive processes. Subjects low in one skill tend to be low in the other skill but not necessarily vice versa because of the asymmetry. Proxy effects also operate, though, because highly fluent readers might have better intellectual skills and previous literacy experiences than low fluent readers. The apparently simple positive correlation is thus confounded by the codependency between skills, the asymmetrical relation between necessary and sufficient skills, and the multicollinearity of multiple factors that affect fluency and comprehension.

A subtle problem concerns the degree of variance within each measure and in the covariance matrix of the two measures. (I assume that the comprehension measure is normally distributed and interacts equally with skewed data to simplify the argument here.) Fluency measures are highly skewed so the variance in fluency measures is usually greatest among those who have low scores on accuracy, rate, and prosody. This will affect the covariance of the two outcome measures in different ways depending on the distribution of scores in the particular sample. Covariance is also affected in subtle ways because of the logical necessity of covariation when fluency and comprehension are measured on the same passage. Regardless of whether fluency and comprehension are measured on the same or separate reading tasks, though, the *lack* of fluency carries the most variance in the correlations, and it may be a proxy for low scores on a host of factors. Thus, "dis-fluency" is correlated with poor reading comprehension, but high fluency is not necessarily correlated with high comprehension.

Techniques for circumventing constraints

Researchers have been aware that reading skills can yield skewed data distributions, but this has usually been regarded as a methodological obstacle and not an underlying conceptual problem. The usual solutions have been to avoid or correct the skewed data rather than to reconceptualize the developing skills or analyze them with nonparametric statistics. It is useful to identify the techniques used in the past to circumvent skewed data and to acknowledge that they are inappropriate. Four techniques will be described.

Subject selection

Most constrained reading skills are acquired in a relatively short developmental time frame. For example, most children learn the letters of the alphabet between 4–7 years of age (or during the first years of formal schooling), and the time for an individual child to master the alphabet is usually less than two years. The age period of mastery reflects both maturational and cultural pressures, so there appears to be a normative developmental period of mastery, but it is possible for mastery to occur later in development or not at all in some cases. The normative periods of skill mastery do not imply that the skills are normally distributed variables or maturational imperatives as much as they imply that the environmental press is normative. Variations in degree of mastery can be compared within and among individuals, but the mistake of traditional research has been to treat variations in the raw performance measures as sets of normally distributed variables.

The most frequently used technique to avoid skewed data on constrained skills is to select subjects for study who are midway to mastery so few individuals are included who are at floor or ceiling levels. This technique yields data that are less skewed and perhaps normally distributed, but the data are special cases. The data do not represent the construct fairly and pose as individual difference variables when they are only temporary differences. Subject selection to achieve normally distributed data is such a common practice that researchers have neglected to question it, but in the case of constrained reading skills, it yields data that are not reliable and interpretations that are misleading.

Alternative measures

The usual measures of developmental growth or achievement are *y*-intercepts that signal acquisition of rules, concepts, skills, and so forth. They allow comparisons between time points within individuals as well as between individuals at the same time. However, there are alternative measures. The course of development of constrained and unconstrained reading skills can be assessed to permit other measures such as

- age of onset and initial skill acquisition (*x*-axis),
- threshold of skill use that signals rapid learning (change in slope),
- length of time (or slope) during the phase of rapid learning,

- point on the curve where mastery is obtained (asymptote, intercept, and slope change), and
- time (or slope) to attain automatic use of the mastered skill.

However, some of these measures may be more appropriate than others to apply to constrained skills. For example, rate of mastery, defined as the age span between onset and asymptote, can be a useful measure, whereas rate of learning, defined as the slope between two points on the growth curve, is not legitimate for constrained skills. It is a poor measure because rate of learning is not linear, and it will vary widely depending on the points assessed in the curve. Rate may appear slow in the beginning and ending phases of the curve and rapid in the middle so rate will be confounded with degree of mastery. Thus, it is possible for a child to exhibit a more rapid rate of learning when mastering the middle 10 letters in his or her 26-letter acquisition than another child who, at the same time, mastered the first or last 5 letters of the alphabet. Of course, the rates within and between children vary widely, but the point is that the rate of learning is not linear, and it is constrained in set size and time so it is a confounded and inappropriate measure with constrained skills. Age of onset and time to master a constrained skill may be more appropriate than rate of learning.

In contrast, unconstrained skills develop continuously, or at least for many more years than constrained skills, and they may vary in proficiency (i.e., intercept measures) across people. They may also exhibit nonlinear differences in slope and rate measures when special experiences promote growth (e.g., intensive instruction or increased opportunities for learning). Unconstrained skills may also depend on precursors in maturation and learning. They may exhibit similar age ranges of rapid growth due to common experiences such as schooling. They may also appear to reach asymptote if learning is not continued, or at least rapid skill growth may slow down after childhood or educational experiences. However, the course and duration of learning are potentially endless for unconstrained skills such as vocabulary and comprehension. This means that the derived research variables are normally distributed, barring measurement constraints that artificially limit the variables.

Task difficulty and complexity

Assessments of reading skills depend on the task difficulty so researchers are careful about matching the abilities of subjects to the task difficulty. That is one reason most assessments use prototypical

exemplars of letters, phonemes, and words when the set of potential elements is sampled. Assessments of letter knowledge using *x*, *q*, and *w* are more difficult than recognizing *a*, *b*, and *c*, so prototypical exemplars are used most often. However, researchers can increase task difficulty by choosing less representative exemplars or including more elements in the task. Phonemic awareness can be assessed with rhyme supply or rhyme choice for 3- to 4-year-olds, but because most 7- to 8-year-olds reach ceiling levels on such tasks, researchers usually give older children harder tasks, such as phoneme elision. Likewise, blending and segmenting phonemes in words may be too easy and may not yield normally distributed data as well as blending and segmenting nonwords. Thus, assessments often include nonwords to rule out familiarity and increase the task difficulty for older children. Manipulating task difficulty to fit the expertise of the sample subjects can manufacture normally distributed data for different samples, but the technique contradicts the construct validity of constrained skills.

Task difficulty can be manipulated and mastery can be made more difficult (even impossible) when assessments of constrained skills are transformed and confounded. For example, rapid letter naming is more difficult than letter recognition, and it may yield a normal distribution of response times, even when the underlying skill of letter recognition is constrained by asymptotic performance. That is because processing speed is an individual difference variable that confounds letter naming with speed of responding, and children's response speeds are variable for many reasons unrelated to reading skills. Decoding nonsense words may likewise appear to yield normally distributed data because of the unfamiliarity and difficulty of the task, when the underlying phoneme recognition is a more constrained skill. Similar changes in task difficulty of most decoding skills can be achieved by manipulating the response demands (i.e., increasing demands on memory or for speedy responses). This technique results in confounded assessments of constrained skills, but it allows researchers to use the tasks with a wider age range without encountering ceiling performance levels, and it creates more normally distributed data than skewed measures of constrained skills. These methods of creating spuriously normal distributions must be identified as inappropriate assessments of constrained skills.

Data transformations

Conceptual constraints in skills can be obscured in the statistical analyses of reading skills. One method is to apply mathematical transforma-

tions to normalize the data when in fact the underlying distribution is not normal and should not be treated as normal. For example, some researchers use simple logarithmic transformations on early literacy predictors (e.g., Lonigan, Burgess, & Anthony, 2000). The reflected square and the reflected logarithm are recommended for transforming data with a strong negative skew (Tabachnick & Fidell, 2001). These processes, along with others such as arc sine transformations, may or may not influence the distribution much, depending on the degree of skew, the variance, and the size of the sample. Rarely do researchers report the degree of correction in skew afforded by data transformations. Never is it acknowledged that such transformations may not be legitimate with constrained skills that are skewed developmentally (e.g., letter knowledge) or methodologically (e.g., accuracy).

A second method is to aggregate variables derived from constrained and unconstrained skills to create composite scores (Riddle-Buly & Valencia, 2002; Wagner, Torgeson, & Rashotte, 1999). Factor analyses and Hierarchical Linear Modeling (HLM) that confound constrained and unconstrained reading skills can produce artificially normal and deceptive variables (e.g., Hecht, Burgess, Torgeson, Wagner, & Rashotte, 2000; Lonigan et al., 2000). When composite scores are created based on factor analyses, HLM, or Item Response Theory (IRT), the result may be an artificially normalized distribution that is more influenced by scores on one measure than another. This problem is exacerbated when researchers aggregate data from highly constrained and less constrained variables. It is also evident when the data are transformed to normalize the distributions, when the sample size is so large that it includes many subjects with floor or ceiling effects, and when the skewed data are blocked to create categorical data (e.g., as in IRT procedures). For example, the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS—K) confounds early reading measures in aggregated variables that obscure developmental differences among the component knowledge and skills (National Center for Education Statistics, 2000). These problems are rarely acknowledged or identified in aggregated or composite measures.

Misinterpreting developmental relations among reading skills

The conditions for misinterpretation of reading data are evident when differences between constrained

and unconstrained skills are revealed. In this section, I discuss four critical misinterpretations of reading research that have implications for educational policies and practices. The first problem is the *transitory correlations* of many basic skills during early development with later achievement measures. The second problem is misinterpretation of simple correlations among reading skills due to multiple correlations with other variables—a *proxy effect*. The third problem is the danger of causal attributions to correlated skills and *prescriptive errors* for instruction. The fourth problem is *transient effects* of interventions that accelerate universally mastered skills. These problems are surprisingly frequent in reading research.

Transitory effects

The clearest illustration of a transitory correlational effect is the relation between children's knowledge of the alphabet and subsequent reading development. For example, Lonigan et al. (2000) said, "knowledge of the alphabet (i.e., knowing the names of letters and the sounds they represent) at entry into school is one of the strongest single predictors of short- and long-term success in learning to read" (p. 597). However, this predictor is highly unstable and very transitory. Careful inspection of longitudinal data reveals decreasing strength of correlations between letter knowledge and later reading proficiency with increasing age, so the predictive validity is short-lived (e.g., Johnston, Anderson, & Holligan, 1996; Muter, 1994). Knowledge of the alphabet is correlated strongly with subsequent reading achievement only during a period of acquisition when the alphabet is partially learned. When measured early or late in the child's acquisition of letter knowledge, the data reveal floor or ceiling effects and decreased correlations with other variables (McBride-Chang, 1999).

Similar patterns of correlations have been observed for other reading skills such as phonemic awareness. Researchers have been aware of the transitory strength of correlations between constrained skills and reading achievement. Wagner et al. (1997) noted in their five-year longitudinal study of early reading predictors that, "Our results suggest that the influence of individual differences in phonological processing abilities on subsequent reading skills is developmentally limited for (letter) naming and is less so for phonological awareness" (p. 477). Walsh, Price, and Gillingham (1988) suggested explicitly that letter knowledge is related to reading achievement in a transitory fashion in their "diminishing returns hypothesis" that shows decreasing correlations as letter

knowledge is mastered. Adams (1990) suggested that the contribution of any particular skill to reading achievement depends on its level of development at the time of testing. Stanovich (2000) said that patterns of correlations among reading skills range from strong to nil depending on the level of skill expertise.

Despite consensus from researchers about the changing patterns of correlations with age, though, many researchers and policymakers erroneously attribute great significance to predictive validity of transitory correlations such as alphabet knowledge. Unfortunately, the well-documented transitory correlations have neither illuminated the distinctive features of constrained skills nor been incorporated in theories of reading development. The constrained skills view explains the transitory effects as a consequence of brief periods of skill mastery followed by continuously reduced variance in the skills as they are universally mastered. It is a narrow developmental period and an empirically special case when the skill is partially acquired by most children that the variance in the sample approximates a normal distribution and can yield strong correlations. Many skills relevant to early reading are mastered in childhood and thus yield asymptotic performance with minimal variance before and after their relatively brief periods of learning. The transitory correlations between mastered skills and other measures of reading have been used as scientific evidence without regard for the lack of significant correlations among the same skills in the same individuals a year later. Transitory correlations are obtained only with constrained skills that display asymptotic levels of performance and nonlinear growth rates. Thus, correlational analyses are inappropriate for distributions derived from constrained skills. New methods of statistical analyses are needed.

Proxy effects

Lonigan et al. (2000) claimed that alphabet knowledge is the best predictor of reading development, but there is an important difference between a predictor and a cause that has been confused. For example, Hecht et al. (2000) noted that composite scores for socioeconomic status (SES) and print knowledge were correlated at $r = .41$. In their longitudinal analyses they observed that the effects of SES on reading achievement were attenuated by 30 to 50% when print knowledge scores were controlled. Therefore, Hecht et al. concluded that "most of the SES related variance in growth of reading skills was accounted for by beginning kindergarten levels of print knowledge" (p. 119). These results led the authors to conclude that,

A practical consequence of the present results is that measures of reading related abilities should be included in test batteries used to identify beginning kindergarten children, particularly those from lower social class backgrounds, at risk for later reading failure.... In addition, the results suggest that preschool and kindergarten interventions involving intensive training in print knowledge, phonological awareness, and/or rate of access skills may help reduce the incidence of later reading failure among children from lower SES families.... (p. 122)

The suggestions for assessment and instruction are causal inferences from correlational data, and similar erroneous reasoning is prevalent in many educational policies. It may be more plausible to interpret print knowledge scores at kindergarten as indirect measures of parental assistance and involvement in helping their children learn to read. Those kindergarten children who scored highly on print concepts, letter names, and letter sounds were most likely to have had more social supports and opportunities for reading, learning, and education than those kindergarten children who scored lower on these tasks. That should be expected by the strong correlation with SES and might be evident if other correlational data were available, such as preschool experiences, parental education, amount of parent-child interactions, or quality of children's literacy materials in the home. Therefore, knowing letter names in kindergarten is probably *not* the mediator of reading comprehension at grades 3 and 4. Instead, home environment and many variables associated with parent-child interactions probably account for better reading comprehension in later grades, especially if those same family and home factors continue to be influential several years later.

The proxy effect is evident for many reading skills because researchers often report bivariate correlations, and exclude analyses of other correlated skills, to show the strength of the correlations for concurrent and predictive validity of early skills. Thus, children's skills on assessments of phonemic awareness, alphabet knowledge, oral reading rate and fluency in grades K-1 have all been shown to be highly correlated with concurrent and future measures of reading achievement, but they are all inter-correlated with one another and many environmental variables. That is why these three classes of skills are among the five "big ideas" about reading that are the foundation of the No Child Left Behind (2002) legislation. That is also why these basic enabling skills are featured in early reading assessment batteries and reading instruction programs. Vocabulary and comprehension, in contrast, have received less attention in early reading assessments.

The canonical correlations have been used to justify the developmental importance and priority of early skills, and the correlations are rarely qualified or interpreted in the context of other factors such as simultaneous developmental accomplishments or social conditions.

It is surprising that so much reading research has followed this model when educational and psychological researchers routinely emphasize the dangers of attributing causality to correlations and of interpreting simple correlations in isolation. Consider this example from the Examiner's Manual of the CTOPP.

On the basis of a growing number of longitudinal correlational and training studies of beginning readers, one may conclude that causal relations seem to exist between the development of phonological processing abilities and the acquisition of word-level reading skills: Some phonological processing skills exert a substantial causal influence on the subsequent development of word reading skills; conversely, letter knowledge exerts a more modest causal influence on the subsequent development of phonological processing abilities.... (Wagner et al., 1999, p. 2)

Wagner et al. (1999) took the causal argument further by claiming that, "A deficit in some aspect of phonological processing ability is viewed as a cause of the most common form of reading disability" (p. 2). This strong claim ignores the mastered components of phonological awareness, the similarity of phonological processing among older children and adults, the transitory nature of the correlations, and the proxy effects of multicorrelated skills during beginning reading. The argument proceeds from correlation to cause to deficit to prescriptions for assessment and instruction of phonological processing skills. Given the problems with transitory correlations among constrained skills, the claims seem overstated.

Prescriptive errors

The proxy effect problem can be minimized if the simple correlations are reinterpreted as indicators and not causes of future reading difficulties. In this interpretation, the multicollinearity among many variables is acknowledged, but three other problems remain. First, some researchers have claimed that children should be taught constrained skills first and often, rather than other reading skills, because measures of alphabet knowledge, print concepts, and phonemic awareness identify children who later read well or poorly. This instructional prescription error is based on a causal attribution to a correlation, and it

ignores other possible variables that contribute to the correlation. The error has been made with instructional prescriptions that emphasize isolated skills such as alphabet knowledge, phonemic awareness, reading rate, and word recognition. Interventions on these skills may facilitate reading development because the skills are important, but, ironically, the instruction may also provide ancillary benefits for practice on related skills, motivation, and guided reading. In other words, the assessment and instruction of constrained skills may be confounded with multiple variables, and they may be effective for reasons other than a causal relation. The argument is analogous to narrow emphases on the benefits of specific vitamins and diets for longevity or weight loss. Both are appealing because they promise simple cures for complex problems, but both are scientifically flawed.

A second problem is that proxy effects are disguised in designs between subjects that confound other differences between groups. Consider reading rate. The correlational evidence that faster readers are better comprehenders is derived from data collected between subjects and between tasks. There is no within-subject evidence that making a specific individual read more quickly engenders better comprehension. For example, teaching children to identify words more rapidly can improve reading rate, but it does not improve text comprehension (Fletcher, Jenkins, & Pany, 1979/1980). The appropriate research is *not* an experimental study between groups of children (one instructed to read fast and one not), but rather the experiment should manipulate reading rate within each reader. Perhaps the study has not been done because the results are so obvious. Forcing faster reading within subjects would not help comprehension at all. Indeed, it seems likely that comprehension would be degraded considerably if a person were forced to read the words more quickly because little attention or processing capacity would be available for understanding. Furthermore, there are advantages to reading slowly such as opportunities for looking back through text, pausing to paraphrase or think about the meaning, and engaging similar strategies to look ahead and back in text to use context to construct meaning. Thus, the prescription to help children read faster is not only a misinterpretation of correlational evidence between subjects; it is counterproductive instruction because it inhibits the use of good comprehension strategies.

A third problem contributing to proxy effects is that the correlational evidence is often gathered between different tasks. For example, researchers correlate oral reading fluency in one-minute samples

with reading comprehension or achievement collected in a different task (Good, Simmons, & Kame'enui, 2001). The interpretive error is that fluency and comprehension are treated as separate, independent, and stable abilities between individuals in the correlations across tasks, yet rate and understanding are necessarily dependent skills within individuals and tasks (Stahl & Hiebert, in press). The commonly accepted speed-accuracy trade-off in reading makes it more surprising that fluent oral reading and comprehension would be considered independent in analyses that correlate the skills on different tasks. The correlations with different tasks between subjects confounds many variables so there are a host of reasons why faster readers might be better comprehenders (i.e., the proxy effect again) at times, and why faster readers should be poorer comprehenders when oral reading speed exceeds the comprehension processing abilities of the reader. Even if rate and comprehension are measured in the same people but on different tasks, there is a presumed and erroneous independence between the skills that will make the correlation difficult to interpret. For example, I might read passage A quickly and have poor comprehension on passage B because it is an unfamiliar topic. The ensuing low correlation between rate on A and understanding on B is an inappropriate measure of the relation between the two skills because rate and comprehension are not individual abilities that are independent of the text, task, and each other. Thus, there are both between-tasks and between-Ss confounds with interpretations of correlations collected in many studies of reading rate and other constrained skills.

Temporary acceleration of mastered skills

The misinterpretations described previously in this article all involve correlational evidence, and some might argue that I am simply replaying an old argument that correlations do not imply causal relations. That is only the superficial conclusion. The deeper issue is that constrained skills have predictably unstable data distributions as readers progress from novices to experts, and those inherently nonnormal distributions cannot and should not be analyzed with parametric statistics, including Pearson correlations. Advocates for the primacy of constrained skills, such as phonemic awareness, in reading development usually cite experimental evidence to bolster the correlational evidence. Studies have shown that direct intervention on alphabet

knowledge, phonics, phonemic awareness, and other constrained skills can increase children's skill levels as well as performance on related tasks such as word identification. For example, Foorman, Fletcher, Francis, Schatschneider, and Mehta (1998) taught first- and second-grade children direct code instruction in letter-sound correspondence and found improvement in phonics. Children also improved their letter-word identification and pseudoword decoding but not their spelling or comprehension. Two conclusions are critical. First, the benefits of intervention were temporary halo effects that show acceleration of normally learned skills. Second, the benefits were evident only on skills similar to the training; there was no transfer to other (unconstrained) reading skills.

The general problems with experimental evidence on constrained skills can be summarized in two ways. First, failure to consider the constraints on mastered skills and knowledge has led researchers to treat constrained and unconstrained variables in the same manner in experimental designs. This error confuses temporary differences in levels of knowledge and proficiency during periods of partial mastery with enduring and stable individual differences that persist over time. It has led to trivial demonstrations that the rates of acquisition of children's constrained skills can be accelerated by special interventions in brief time periods. However, the effects are transient, they fail to generalize to other reading skills and knowledge, and they fail to lead to any long-term advantages in reading achievement. This is similar to other developmental research that has shown temporary gains in learning skills such as conservation that are universally acquired. They are ephemeral and transitory because the skills are universally mastered.

A second point about temporary instructional effects is the dependency on developmental timing of the interventions. The effectiveness of treatments must be analyzed for individuals, not groups, in order to understand the developmental appropriateness of the interventions. It seems reasonable to speculate that the most effective treatments occur at the children's leading edges of their "zones of potential development." This means that treatments vary in their appropriateness and effectiveness according to the age and expertise of the child. Interventions must be tailored to individual growth trajectories so group instruction on universally mastered skills and knowledge is unlikely to be effective unless most members of the group lack the instructed knowledge and skills. The success of direct instructional methods, such as the curriculum used in the Foorman et al.

(1998) study, is most likely in classrooms where all children lack universally mastered skills and knowledge and instruction provides a temporary boost to constrained skills. The risks are evident in the unscientific and sweeping claims made about the effectiveness of the instruction for long-term reading development and achievement.

I hasten to add that this view of halo effects does not mean that instruction on knowledge and skills involved in early reading development should be withheld from children. On the contrary, 4- to 7-year-olds need to be instructed on those skills early and persistently by teachers and parents. Constrained skills must be mastered. What is unscientific, illogical, and unwarranted are the claims that one kind of instruction is the best or only way to promote the acquisition of the skills, that those methods are uniformly appropriate for all children, that the instructed skills have greater priority than other skills, and that such interventions prevent reading failure.

Conclusions and implications

The major point of this article is that reading researchers have neglected different conceptual, developmental, and methodological constraints on skill trajectories, and thus they have treated all reading skills as similar components compiled in expert reading. My claim is that some reading skills are universally mastered in a relatively brief developmental time span, and thus they should not be regarded conceptually or methodologically as normally distributed skills. These skills are only normally distributed in selected samples of children who are midway to mastery, and this distribution is transitory and artificial. Consequently, the characteristics of the sample will determine the degree of mastery that in turn determines the degree of skew in the data that in turn influences the degree of correlation found between the skills and other variables. It should not be surprising that different studies with different samples of children find different patterns of correlations with constrained skills. It also should not be surprising that different studies repeating the same sampling errors and same data transformation errors can corroborate similar correlations. These results can be explained parsimoniously by the constrained skills view. Reexamination of the skew and variance in relation to mastery levels of constrained skills can explain when high or low correlations are obtained.

The distinction between constrained and unconstrained skills has important implications for both theories of reading development and assessments

of reading development. Theories of reading development may benefit by distinguishing among constrained and unconstrained skills because of their different scope and developmental trajectories. Although it is important for children to master constrained skills, the advantages of mastery at an early age (or a rapid rate of mastery) may be important only for a limited time. Likewise, the advantages may be quite specific to the constrained skills. Thus, rapid mastery of constrained skills may have transitory importance, limited scope, and limited range of influence on enduring reading achievement.

Some researchers argue that the evidence from training studies shows the causal influences of constrained skills on reading proficiency. Although constrained skills are necessary for many reading skills (e.g., comprehension depends in part on skilled decoding), they are not sufficient independent causes of reading development. Experimental evidence that short-term training can increase children's letter knowledge, concepts of print, or phonemic awareness does not establish a causal role in development because those effects are small and temporary (Adams, 1990; Muter, 1994). In a review of experimental training studies on phonological awareness, Bus and van IJzendoorn (1999) found that only 12% of the variance in reading is accounted for by phonological processing skills. (However, this estimate is entirely dependent on the level of mastery of the specific sample.) They suggested that the small effects might be due to decay over time or ceiling performance on the tasks—that is, mastery of constrained skills.

I do not want to restart the reading wars, but there is simply no valid and reliable scientific evidence that indicates that the relative degree of mastery of constrained skills is the cause of later reading proficiency or comprehension, even if it is a good predictor of future proficiency. Lack of constrained skills, dis-fluency, and lack of basic knowledge about letters, phonological awareness, and concepts about print are associated with lack of reading achievement, for many important reasons that deserve attention. However, there are other cognitive, social, motivational, and linguistic factors involved in reading proficiency that are important causal factors in reading development. Perhaps mastery of constrained skills at some threshold levels enables other skills to develop, but once mastered, the constrained skills are insufficient to predict or promote future reading development.

Many claims of early reading assessments appear unwarranted in light of the new view of constrained skills. The predictive validity of early reading assessments is mostly due to multicollinearity of data

that shows that early proficiency of reading skills is correlated with later reading achievement. This correlation depends on the particular sample of children and their relative skill expertise, and the relations are transitory at best and deceiving at worst. The correlations are due to between-Ss differences, so the statistics are proxies for differences that might be due to many experiential factors that distinguish early rapid learners from children who acquire the mastered skills more slowly. On the positive side, it is important to note that constrained skills need to be mastered because they are necessary but not sufficient for other reading skills. They enable automatic decoding, deployment of attention, and application of comprehension strategies so they set the stage for reading development, but they are not simple causes for complex reading skills to develop. Indeed, unconstrained skills such as vocabulary and comprehension develop before, during, and after constrained skills are mastered so there is no evidence to warrant instructional priority of constrained skills over unconstrained skills.

The implication for assessments of reading development is that the data derived from assessments of constrained skills may exhibit transitory and unstable relations with other variables, especially when they are analyzed by parametric statistics. All claims about the predictive and concurrent validity of basic skills, such as phonological awareness, letter-sound correspondence, letter identification, phonics, and concepts about print, need to be reinterpreted if they are based on correlations of data from constrained skills. Likewise, all data derived from constrained skills that are analyzed with factor analyses, cluster analyses, HLM, and path analyses are also suspect because the analyses embed constrained skills in composite measures that confound relative levels of mastery, different degrees of skew in the data, and potential floor and ceiling effects. When the underlying construct is not normally distributed, it is inappropriate to use parametric statistics for analyses, even if careful sampling and transformations reduce skew in the data. This is a sweeping and controversial indictment of traditional data analyses in reading research because many studies have used parametric statistics with constrained skills to establish validity of assessment instruments. The solutions require extensive discussion but should include nonparametric statistical methods such as conditional probability, contingency tables, log linear models, and other techniques for analyzing constrained skills in a categorical manner.

On a practical level, educators should be wary of policies that require repeated assessments of

constrained skills as indicators of (a) individual reading achievement or (b) successful programs. One danger is that excessive testing of constrained skills may lead to an overemphasis on these skills to the exclusion of unconstrained skills such as vocabulary and comprehension. A second risk is that policymakers and the public may equate success on constrained skills with reading proficiency. This would create a minimum competency approach to reading assessment that does not adequately assess children's emerging use and control of literacy. The recent proliferation of state-designed assessments of early literacy and mandated policies for repeated testing of basic skills in elementary grades may exacerbate these risks.

What does this mean for policymakers? First, it suggests that most of the scientific evidence about reading skills and reading development, particularly related to decoding skills, is based on inadequate theories, measures, and interpretations. The constrained skills view illuminates these problems. Second, it suggests that the criteria and methods used to identify valid assessments of early reading should be reexamined. Third, new theories of reading skills and reading development, along with new methods and measures, need to be created and tested in order to provide reliable and valid assessments of reading development for children. Much more scientific research needs to be conducted before sweeping policies are legislated. Meanwhile, teachers need to be provided with useful and multidimensional assessments of children's reading skills that are designed to support their diagnostic evaluations and instructional decisions about individual children. Anything less is a disservice to teachers and students.

SCOTT G. PARIS is a professor of psychology and education at the University of Michigan and is currently chair of the graduate program in psychology and a principal investigator at the Center for the Improvement of Early Reading Achievement. Paris has published psychology textbooks, has contributed to several basal reading series, and has published research in the areas of children's reading, learning, memory, metacognition, and cognitive development. He can be contacted at the University of Michigan, Department of Psychology, 2008 East Hall, Ann Arbor, MI 48109, USA, or by e-mail at sparis@umich.edu.

REFERENCES

- ADAMS, M.J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- ADAMS, M.J., & TREADWAY, J. (2000). *The fox in the box*. Monterey, CA: CIB/McGraw-Hill.
- BERLINER, D.C. (2002). Educational research: The hardest science of all. *Educational Researcher*, 31(8), 18–20.
- BURGESS, S.R., HECHT, S.A., & LONIGAN, C.J. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly*, 37, 408–426. doi:10.1598/RRQ.37.4.4
- BUS, A.G., & VAN IJZENDOORN, M.H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403–414.
- CHALL, J.S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.
- CHALL, J.S. (1996). *Stages of reading development* (2nd ed.). Orlando, FL: Harcourt Brace.
- CLAY, M.M. (1979). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.
- CLAY, M.M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.
- EHRI, L.C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, 18, 116–125.
- FEUER, M.J., TOWNE, L., & SHAVELSON, R.J. (2002). Scientific culture and educational research. *Educational Researcher*, 31(8), 4–14.
- FLEISHER, L.S., JENKINS, J.R., & PANY, D. (1979/1980). Effects on poor readers' comprehension of training in rapid decoding. *Reading Research Quarterly*, 15, 30–48.
- FOORMAN, B.R., FLETCHER, J.M., FRANCIS, D.J., SCHATSCHNEIDER, C., & MEHTA, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, 90(1), 37–55.
- GOOD, R.H., & KAMINSKI, R.A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
- GOOD, R.H., SIMMONS, D.C., & KAME'ENUI, E.J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5(3), 257–288.
- HASBROUCK, J.E., & TINDAL, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children*, 24(3), 41–44.
- HECHT, S.A., BURGESS, S.R., TORGESON, J.K., WAGNER, R.K., & RASHOTTE, C.A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing: An Interdisciplinary Journal*, 12, 99–127.
- JOHNSTON, R.S., ANDERSON, M., & HOLLIGAN, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in pre-readers: The nature of the relationship. *Reading and Writing: An Interdisciplinary Journal*, 8, 217–234.
- KIBBY, M.W. (1979). Passage readability affects the oral reading strategies of disabled readers. *The Reading Teacher*, 32, 390–396.
- KUHN, M.R., & STAHL, S.A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95(1), 3–21.
- LABERGE, D., & SAMUELS, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- LIPSON, M.Y., & WIXSON, K.K. (2003). *Assessment and instruction of reading and writing difficulty*. Boston: Allyn & Bacon.
- LONIGAN, C.J., BURGESS, S.R., & ANTHONY, J.L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology*, 36(5), 596–613.
- MCCBRIDE-CHANG, C. (1999). The ABCs of the ABCs: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly*, 45(2), 285–308.
- MORRIS, D., BLOODGOOD, J.W., LOMAX, R.G., & PERNEY, J. (2003). Developmental steps in learning to read: A longitudinal study in kindergarten and first grade. *Reading Research Quarterly*, 38, 302–328. doi:10.1598/RRQ.38.3.1
- MUTER, V. (1994). Influence of phonological awareness and letter knowledge on beginning reading and spelling development. In C. Hulme (Ed.), *Reading development and dyslexia* (pp. 45–62). London: Whurr Publishers.
- NATIONAL CENTER FOR EDUCATION STATISTICS. (2000). *Early childhood longitudinal study, kindergarten class of 1998–99*. Washington, DC: U.S. Department of Education.
- NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
- NO CHILD LEFT BEHIND ACT OF 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

- PARIS, S.G., CARPENTER, R.D., PARIS, A.H., & HAMILTON, E.E. (in press). Spurious and genuine correlates of children's reading comprehension. In S.G. Paris & S.A. Stahl (Eds.), *Current issues in reading comprehension and assessment*. Mahwah, NJ: Erlbaum.
- PARIS, S.G., & PARIS, A.H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, 36(2), 89-101.
- RIDDLE-BULY, M., & VALENCIA, S.W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, 24, 219-239.
- RUMELHART, D.E. (1994). Toward an interactive model of reading. In R.B. Ruddell, M. Rapp-Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 864-894). Newark, DE: International Reading Association.
- SNOW, C.E., BURNS, M.S., & GRIFFIN, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- STAHL, S.A., & HIEBERT, E. (in press). The assessment of fluency and other component skills and knowledge: Contributions to comprehension. In S.G. Paris & S.A. Stahl (Eds.), *Current issues in reading comprehension and assessment*. Mahwah, NJ: Erlbaum.
- STANOVICH, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.
- STANOVICH, K.E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford.
- TABACHNICK, B., & FIDELL, L. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.
- TEXAS EDUCATION AGENCY. (2001/2002). *Texas Primary Reading Inventory (TPRI)*. Austin, TX: Author.
- THOMPSON, A.C., & BINDER, K.S. (2003). A comparison of the factors affecting reading performance of functionally illiterate adults and children matched by reading level. *Reading Research Quarterly*, 38, 236-258.
- WAGNER, R.K., TORGESON, J.K., & RASHOTTE, C.A. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.
- WAGNER, R.K., TORGESON, J.K., RASHOTTE, C.A., HECHT, S.A., BARKER, T.A., BURGESS, S.R., ET AL. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, 33, 468-479.
- WALSH, D.J., PRICE, G.G., & GILLINGHAM, M.G. (1988). The critical but transitory importance of letter naming. *Reading Research Quarterly*, 23, 108-122.

Received March 12, 2003

Final revision received May 24, 2004

Accepted June 29, 2004

